Image manifolds which are isometric to Euclidean space

David L. Donoho and Carrie Grimes *†

Abstract

Recently, the Isomap procedure [1] was proposed as a new way to recover a low-dimensional parametrization of data lying on a low-dimensional submanifold in high-dimensional space. The method assumes that the submanifold, viewed as a Riemannian submanifold of the ambient high-dimensional space, is isometric to a convex subset of Euclidean space. This naturally raises the question: what datasets can reasonably be modeled by this condition? In this paper, we consider a special kind of image data: families of images generated by articulation of one or several objects in a scene – for example, images of a black disk on a white background with center placed at a range of locations. The collection of all images in such an articulation family, as the parameters of the articulation vary, makes up an articulation manifold, a submanifold of L^2. We study the properties of such articulation manifolds, in particular, their lack of differentiability when the images have edges. Under these conditions, we show that there exists a natural renormalization of geodesic distance which yields a well-defined metric. We exhibit a list of articulation models where the corresponding manifold equipped with this new metric is indeed isometric to a convex subset of Euclidean space. Examples include translations of a symmetric object, rotations of a closed set, articulations of a horizon, and expressions of a cartoon face.

The theoretical predictions from our study are borne out by empirical experiments with published Isomap code. We also note that in the case where several components of the image articulate independently, isometry may fail; for example, with several disks in an image avoiding contact, the underlying Riemannian manifold is locally isometric to an open, connected, but not convex subset of Euclidean space. Such a situation matches the assumptions of our recently-proposed Hessian Eigenmaps procedure, but not the original Isomap procedure.

Keywords: Isomap; Multidimensional Scaling; Manifolds of Articulated Images; Isometry.

Acknowledgments: CG would like to thank Vin de Silva and DLD would like to thank Raphy Coifman, in both cases for discussions and references. This work has been partially supported by National Science Foundation grants DMS-00-72661, ANI-00-85984, and DMS-01-40698.

1 Introduction

An object in the world can be imaged in various ways: it can be observed from various distances and orientations; also the object itself can change appearance in pose and articulation; and

*Department of Statistics, Stanford University, 390 Serra Mall, Stanford, CA 94305-4065 {donoho,carrie}@stat.stanford.edu

†This work has been partially supported by National Science Foundation grant DMS 00-72661, and by DARPA Applied and Computational Mathematics Program

finally the lighting can vary in spectral balance. The collection of all such images of an object, as the parameters (locations, scale, pose, light color, etc.) vary, can be thought of as a manifold in the high-dimensional space of all conceivable images; compare Nayar et al. [2], Belhumeur and Kriegman [3].

Often one is given many digital images ($I_i : i = 1, \ldots, N$) of the same object in a variety of articulations and poses, but the data are 'unlabeled' in the sense that the underlying parametrization and the corresponding parameter values are unknown. The images are thus thought to arise from a manifold M, without the manifold or the associated parametrization being known. For concrete tasks in image understanding and image coding, it could be useful to 'learn' the structure of such image articulation manifolds and to recover the underlying parameters (location, scale, etc.) from unlabeled data. This could be helpful for recognizing, for example, articulated vehicles in target recognition or faces with different expressions in facial recognition.

The general problem of learning the shape of a manifold from scattered observations has been around for a long time; it has been the source of many multivariate techniques, including principal components analysis, independent components analysis, multidimensional scaling, self-organizing mappings, and other important methodological developments.

Recently, Tenenbaum et al. [1] proposed the Isomap procedure as a general tool for recovering the unknown parametrization underlying a set of digital images, $\{I_i\}$, of faces in various attitudes and articulations. The general principle of Isomap is to measure distance between images, not using Euclidean distance (which obscures the intrinsic manifold structure), but using distance according to the shortest path in a nearest neighbor graph; and to use this graph distance as input to a classical "principal coordinates" multidimensional scaling procedure.

1.1 Validating Isomap

Tenenbaum et al. [1] published a few interesting examples, for example mapping out the parameters underlying a face seen from a variety of viewpoints. These empirical successes lead to the ...

> **Obvious Question:** how "correct" is the Isomap procedure; does it really recover the "true" underlying parametrization of families of articulated images?

This leads immediately to the ...

> **Obvious Approach:** Test Isomap for synthetic data where we know *a priori* "the natural" parametrizations and see if it can recover the parametrization of image

manifolds.

In following the 'obvious' approach, we would construct synthetic datasets of artificial images undergoing standard articulations – translations, rotations, etc. – and ask if the Isomap parametrization correctly discovers the (known) underlying parametrizations.

There are several reasons such a purely empirical investigation based on running Isomap on artificial examples may not be enlightening.

- *Sampling issues.* Whether Isomap works or not might depend on how many model images I_i are in the database. In some vague sense, it will be important for these images to be well-distributed across the image manifold, but exactly what this means in a specific instance is unclear. Consequently, if Isomap 'fails', this may be due to poor data sampling rather than any intrinsic property of Isomap.

- *Digitization issues.* In some sense, the fact that images are discretized into pixels makes them 'noisy' / 'blocky'; this again causes some difficulty in interpreting 'failure' of Isomap – is it due to the pixelization or is it intrinsic to the type of articulation?

- *'Big Picture' issues.* Empirical work really doesn't give us an intellectual framework that we can leverage into other settings, at least not the kind of framework that might be possible by a more theoretical approach.

For these and other reasons, we propose to develop an alternate framework for understanding the behavior of Isomap on image manifolds. We think of an image as a function $I(x)$ of a continuous variable $x \in \mathbb{R}^2$. We consider articulations of a base image I_0, producing a family of images $\{I_\theta : \theta \in \Theta\}$, where θ is the parameter of the articulation and Θ is the parameter space. Examples we will consider include translation families, where $\Theta = \mathbb{R}^2$ and $I_\theta(x) = I_0(x - \theta)$. In this "continuum" viewpoint, neither sampling nor digitization can cause problems, and a clear intellectual framework exists naturally.

We consider families of images as a subset of $L^2(\mathbb{R}^2)$ and, initially, measure distance between images by $\mu(\theta_1, \theta_2) = \|I_{\theta_1} - I_{\theta_2}\|_{L^2}$. The subset $M = \{I_\theta : \theta \in \Theta\}$ becomes a metric space (M, μ) and, one can check, a continuous submanifold of $L^2 = L^2(\mathbb{R}^2)$.

Translating the isometry principle underlying Isomap into this continuum setting, we let $G(\theta_0, \theta_1)$ be the geodesic distance defined by taking the shortest path lying in M going from I_{θ_0} to I_{θ_1}. Then we ask if $G(\theta_0, \theta_1)$ is proportional to Euclidean distance between θ_0 and θ_1, for all pairs (θ_0, θ_1). In short, we ask if (Θ, G) is isometric to $(\Theta, \|\cdot\|)$ up to a constant scaling factor. When the indicated isometry property holds, then it is possible to recover, from knowledge

of M alone, the parameter space Θ and the parametrization $\theta \leftrightarrow I_\theta$, modulo a rigid motion and rescaling of the parameter space. We will therefore also say, when isometry holds, that *Continuum Isomap works*.

Our approach provides a clear theoretical framework with crisply-stated problems that are easily answered. There will, however, be a certain amount of technical detail required to develop this view, having to do with the fact that when images have edges, the underlying 'surface' $M \subset L^2(\mathbb{R}^2)$ is typically not a differentiable manifold; indeed, the geodesic distance G will not be finite. We will derive a renormalized geodesic distance δ, a rescaled limit of distances defined on smoothed versions of M as the smoothing parameter tends to zero. Finally our criterion for success in the continuum setting will be that (Θ, δ) is isometric to $(\Theta, \|\cdot\|)$.

1.2 Results

Our criterion for saying that 'Continuum Isomap works' is quite stringent, demanding exact isometry of the underlying metric spaces, and therefore it may seem unlikely that it could ever hold. As it turns out, in the image articulation setting that we describe, there are a number of interesting image libraries where isometry really does hold. These cases include:

- Translation of simple black objects on a white background;
- Pivoting certain simple black objects on a white background around a fixed point;
- Morphing of boundaries of black objects on a white background;
- Articulation of 'fingers' of a digital 'hand';
- Articulation of a cartoon face by arranging its eyebrows, eyelids, and lips.

It also turns out that isometry holds for certain 'movies', i.e. for images articulating in time; an example being a cartoon face gesturing in time according to a sufficiently rich and complete inventory of gestures. Our theory predicts that from 'watching' sufficiently many movies of a cartoon face gesturing in time, Isomap could in principle will correctly learn the correct parametrization underlying the facial gesturing. Many such results are possible to state, but we do not pursue them in this paper.

Our examples provide a theoretical framework, in the setting of image databases, which validates the claim implicit in early uses of Isomap. However, we also point out examples of image articulation manifolds where isometry fails. At one extreme, local isometry holds, but the underlying parameter space is not convex; this is the case when there are several articulating

objects which avoid each other – the exclusion requirement creates a nonconvex parameter space. In other cases, even local isometry fails, for example, when objects occlude.

A special feature of our approach is the analysis framework we have developed. In this framework we view images as defined on a continuum, and we translate an algorithm of interest into a continuum version of the algorithm. We have applied the framework here to understanding the Isomap procedure, but it could equally well be applied to understanding other manifold learning procedures, such as Locally Linear Embedding (LLE) [4]. We in fact used this framework to understand LLE specifically, and what we learned from the effort led us to define a new procedure, Hessian Eigenmaps [5], introduced in other work, that can solve the local isometry problem defined here. Potentially, the 'continuum framework' could be applied usefully to additional algorithms.

The article is organized as follows: Section 2 develops the mathematical foundations of renormalized distance δ; Section 3 gives a few general 'simple' image families where δ can be calculated; Section 4 tests the theoretical results against generated image data; Section 5 interprets the methodology for more complex example image families (such as a horizon parametrized by a basis function expansion, or composite articulations of more than one object). Section 6 discusses a more elaborate example: movies; Section 7 offers examples where isometry fails; and further issues and discussion appear in Section 8.

2 Geodesic Distance Between Images

In this section, we develop a continuum setting for understanding the issues behind the Isomap procedure. This involves a review of some basic geometric notions, such as arclength and geodesic, showing the ill-posedness of the 'obvious' approach to Continuum Isomap, and defining a special renormalized distance and formulas for computing it.

2.1 A Simple Manifold of Images over the Continuum

Consider the family of smooth functions $f_0 : \mathbb{R}^2 \to \mathbb{R}^2$ that are radially symmetric with unit L^2-norm. For a parameter $\theta \in \mathbb{R}^2$, we then have the translation family

$$f_\theta(x_1, x_2) = f_0(x - \theta), \tag{2.1}$$

so that the 'image' is being shifted. A simple calculation shows that

$$\begin{aligned} \|f_{\theta_1} - f_{\theta_0}\|_{L^2} &= \left(\int |f_{\theta_1}(x) - f_{\theta_0}(x)|^2 \, dx \right)^{1/2} \\ &= \eta(\|\theta_1 - \theta_0\|), \end{aligned} \tag{2.2}$$

where $\eta : \mathbb{R}^+ \to \mathbb{R}$ is a continuous nonlinear function satisfying $\eta(0) = 0$. Moreover, when $\|\theta_0 - \theta_1\|$ is large, f_{θ_0} and f_{θ_1} are almost entirely separated and therefore $\eta(\infty) = \sqrt{2}$.

Suppose $\Theta = \{\|\theta\| \leq 1\}$, and set $M = \{f_\theta : \theta \in \Theta\}$. M is a 2-dimensional subset of the ambient space $L^2(\mathbb{R}^2)$.

Define the metric $\mu(\theta_0, \theta_1) = \|f_{\theta_0} - f_{\theta_1}\|_{L^2(\mathbb{R}^2)}$. The relation $\mu(\theta_0, \theta_1) = \eta(\|\theta_0 - \theta_1\|)$ means that, while the topology of M is equivalent to the Euclidean topology, the submanifold is a curved subset of L^2.

2.2 Geodesic Distance

Let M be a submanifold of L^2, and consider now a curve $\gamma : [0, 1] \to M$. Suppose that γ is C^1-smooth when viewed as a curve $[0, 1] \to L^2(\mathbb{R}^2)$. Thus for each $t \in [0, 1]$ there exists a derivative $\dot{\gamma}(t) : \mathbb{R}^2 \to \mathbb{R}$ that has finite L^2-norm and that varies continuously in t. Then a first definition of the length of γ is

$$L(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|_{L^2} \, dt.$$

A more general definition of arclength, which agrees with the calculus definition when γ is C^1-smooth, sets

$$L(\gamma) = \sup \sum_{i=0}^{n-1} \|\gamma(t_{i+1}) - \gamma(t_i)\|_{L^2},$$

where the supremum is taken over all partitions of the unit interval with arbitrary breakpoints $0 = t_0 < t_1 < ... < t_n = 1$. This definition is valid even for merely continuous curves γ.

The geodesic distance between points θ_0 and θ_1 in the smooth manifold M is the length of the shortest curve joining the two points, or

$$G(\theta_0, \theta_1; M) = \inf\{L(\gamma) : \gamma(0) = f_{\theta_0}, \, \gamma(1) = f_{\theta_1}\}.$$

We note that G is the distance between points in M, when it is viewed as a Riemannian submanifold of the ambient space L^2. As a simple example, suppose f_0 is a smooth radial function

on $\mathbb{R}^2$, such as the Gaussian $f_0(x) = e^{-\frac{\|x\|^2}{2}}$, and $f_\theta(\mathbf{x}) = f_0(\mathbf{x} - \theta)$. Then

$$G(\theta_1, \theta_0; M) = c\|\theta_1 - \theta_0\|_2 \tag{2.3}$$

where

$$c^{-1} = \|\frac{\partial}{\partial x_1} f_0\|_{L^2(\mathbb{R}^2)}.$$

In short, up to a constant, *geodesic distance agrees globally with Euclidean distance on the natural parameter space.* We can then conclude that M, viewed as a Riemannian submanifold of L^2 is isometric to $(\Theta, \|\cdot\|)$.

2.3 Recovering the Natural Parametrization

The idea behind Isomap [1] is that an isometry between M (viewed as a Riemannian submanifold of an ambient space) and a subset of $\mathbb{R}^d$ allows one to recover the underlying isometric parametrization. Isomap exploits the usual procedure of metric multidimensional scaling [6, 7]. Given a matrix of pairwise distances $(d_{i,j} : 1 \leq i, j \leq n)$ between pairs of objects, and supposing these are truly the Euclidean distances between points in a d-dimensional Euclidean space, this procedure returns n points $\mathbf{x}_i$, say, obeying

$$\|\mathbf{x}_i - \mathbf{x}_j\| = d_{i,j}.$$

The Isomap proposal is to use specially-derived empirical geodesic distances for the $d_{i,j}$.

Applying this idea in the continuum image setting, if we use true geodesic distances $d_{i,j} = G(\theta_i, \theta_j; M)$ and the isometry property (2.3) holds, we would obtain a set of points $(\mathbf{x}_i)$ obeying

$$\mathbf{x}_i = U\theta_i + b.$$

for some fixed orthogonal matrix U and vector b.

Definition 2.1 *Suppose that the geodesic distance between points in M, viewed as a submanifold of L^2, is proportional to Euclidean distance in the parameter space Θ. Then we say that (Θ, G) and $(\Theta, \|\cdot\|)$ are in isometry, and that Continuum Isomap works.*

Of course, isometry allows recovery only up to a scaling and rigid motion. Summarizing the above discussion, we have:

Corollary 2.1 *Suppose we have a parametrized family of 'images' $f_\theta : \mathbb{R}^2 \mapsto \mathbb{R}$ defined by translation of a common prototype: $f_\theta(\mathbf{x}) = f_0(\mathbf{x} - \theta)$, that f_0 is radially symmetric, belongs to L^2 and is differentiable in L^2. Then, for an appropriate $c > 0$, the geodesic distance between f_{θ_0} and f_{θ_1} has the form*

$$G(\theta_0, \theta_1) = c \cdot \|\theta_0 - \theta_1\|.$$

Hence isometry holds and Continuum Isomap works.

For complete clarity, we also point out that, in these cases, using ordinary L^2 distance rather than geodesic distance would *not* recover the true parameter space. Indeed as we have seen $\mu(\theta_0, \theta_1) = \eta(\|\theta_0 - \theta_1\|)$ where $\eta(\cdot)$ is nonlinear. Therefore, the properties of geodesic distance make it uniquely successful in this setting.

2.4 Non-differentiable Image Manifolds

The above definitions rely on the implicit assumption that the submanifold M is differentiable, with finite geodesic distances everywhere. Unfortunately, objects in the real world have edges, and so a respectable model of images on the continuum has discontinuities at edges of the object boundaries. This turns out to mean that the submanifold M is typically non-differentiable, and that typically the geodesics do not have finite length.

Consider the indicator of the unit disk,

$$I_0(x) = 1_{\{|x| \leq 1\}},$$

and a family of translated images, $\theta \in \Theta = \mathbb{R}^2$, $I_\theta = I_0(\mathbf{x} - \theta), \theta \in \Theta$. Then let $M = \{I_\theta\}$. Now we can write $\mu(\theta_1, \theta_0; M) = \|I_{\theta_1} - I_{\theta_0}\|_{L^2} = \eta(\|\theta_1 - \theta_0\|)$, for a new function $\eta(\cdot)$. Indeed, $\|I_{\theta_1} - I_{\theta_0}\|$ is just the square root of the area formed by the symmetric difference between the support of I_{θ_1} and I_{θ_0}. This area depends on $r = \|\theta_1 - \theta_0\|$ only. Also, we can calculate that η is not differentiable at the origin: $\eta(d) \asymp d^{1/2}$. Indeed when $r = \|\theta_1 - \theta_0\|$ the area A_r of the symmetric difference $\mathrm{supp}(I_{\theta_0}) \Delta \, \mathrm{supp}(I_{\theta_1})$ obeys $A_r \sim cr$ as $r \to 0$. Hence its square root obeys $\sqrt{A_r} \sim \sqrt{cr}$ as $r \to 0$, which implies the asymptotic equivalence $\eta(d) \asymp d^{1/2}$.

We see immediately that every path in M has infinite length. Indeed, using the extended notion of arclength discussed in Section 2.2, and setting $r_n = \|I_{1/n} - I_0\|_{L^2}$,

$$L(\gamma) \geq \sum_{k=0}^{n-1} \|I_{(k+1)/n} - I_{k/n}\| = n \cdot \eta(r_n) \asymp n \cdot \eta(c/n) \approx n \cdot n^{-1/2} = n^{1/2},$$

which increases without bound as $n \to \infty$. This means that the mapping $\theta \mapsto I_\theta$ cannot be differentiable in L^2.

More generally, we have:

Theorem 2.1 *If M is the articulation manifold gotten by translating an image with edges, then M is a continuous but not a differentiable manifold. Moreover paths in M do not have finite length.*

For the formal proof of this result, see [8]. In short, *geodesic distance is typically not well-defined on an image articulation manifold when the generating image has edges.* Since any reasonably interesting image has edges, this seems at first glance to be fatal to the notion of using isometry to uncover parametrization in the continuum setting.

2.5 Renormalized Geodesic Distance

To rescue the notion of isometry in cases where the underlying image has edges, we introduce a process of regularization and renormalization.

Consider again the case where $I_0 = 1_{\{|x|\leq 1\}}$ is the indicator of the disk, and define the regularized object I_0^h by $I_0^h = I_0 \star \phi_h$ where ϕ is a smooth radial function of width h and $\star$ stands for convolution. Each image $I_\theta^h = I_0^h(\mathbf{x} - \theta)$ is smooth, and if we let M_h denote the image manifold generated by translations of I_θ^h, with L^2 metric $\mu(\theta_1, \theta_0; M_h) = \|I_\theta^h - I_{\theta_0}^h\|_{L^2}$, we get that M_h is a smooth submanifold of L^2. Obviously, M_h tends to M as $h \to 0$, and if we consider a curve $\vartheta(t)$ in the common parameter space Θ of all the M_h and of M, then the lengths of the induced paths on the respective manifolds M_h must converge to the lengths of the limit manifold M. It follows that, if we let $G(\cdot, \cdot; M_h)$ denote the corresponding geodesic distance on M_h, we have

$$\lim_{h \to 0} G(\theta_1, \theta_0; M_h) = \infty.$$

In short, regularization makes the geodesic distance finite, but this effect disappears as $h \to 0$.

To get a finite limiting result, we make the observation that, if τ_0 and τ_1 are two fixed 'landmarks', then the ratio

$$\frac{G(I_{\theta_0}^h, I_{\theta_1}^h; M_h)}{G(I_{\tau_0}^h, I_{\tau_1}^h; M_h)}$$

need not diverge as $h \to 0$. In effect, if we renormalize the distance so that τ_0 and τ_1 stay at unit distance on M_h for all $h > 0$, we get a finite limiting ratio.

In our running example, let I_0 be the indicator of the unit disk in $\mathbb{R}^2$ and let $I_\theta(\mathbf{x}) = I(\mathbf{x} - \theta)$ be translation by θ. For regularization, let ϕ_h denote the standard bivariate Gaussian density

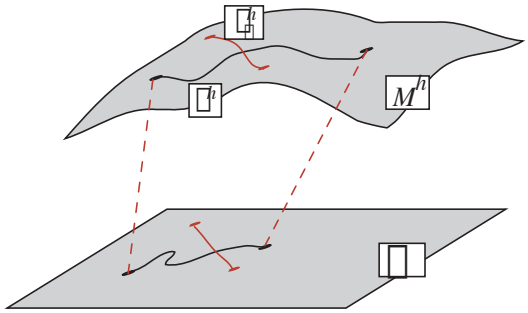

Figure 1: Correspondence between parameter space and smoothed manifold

with covariance matrix $h^2 \cdot Id$. Let the path $\vartheta_\tau(t)$ be the line segment between $\tau_0 = (0,0)$ and $\tau_1 = (1,0)$. The length of the corresponding path (the 'landmark distance'), γ_τ^h, on the smooth manifold M_h generated by $\gamma_\tau^h(t) = I^h_{\vartheta(\tau)}$ is given by

$$L(\gamma_\tau^h) = \int_0^1 \|\frac{\partial}{\partial t}\gamma_\tau^h(t)\|_{L^2(\mathbb{R}^2)}dt = \int_0^1 \|\frac{\partial}{\partial t} I^h_{\vartheta_\tau(t)}\|_{L^2(\mathbb{R}^2)}dt. \tag{2.4}$$

Now note that because of the translation invariance of the L^2 norm, we can write, for a a constant $A_{h,\tau}$ not depending on t,

$$\|\frac{\partial}{\partial t} I^h_{\vartheta_\tau(t)}\|_{L^2(\mathbb{R}^2)} = \|\frac{\partial}{\partial t} I^h_{\vartheta_\tau(t)}\|_{L^2(\mathbb{R}^2)}|_{t=0} = A_{h,\tau},$$

so $L(\gamma_\tau^h) = A_{h,\tau}$. Now for any other chosen parameter points θ_0 and θ_1 in $\mathbb{R}^2$, connected by a line segment, consider the corresponding path γ_{01}^h in M_h between $I^h_{\theta_0}$ and $I^h_{\theta_1}$. We can perform the same calculation and see that

$$L(\gamma_{01}^h) = A_{h,\tau} \cdot \|\theta_0 - \theta_1\|.$$

Moreover γ_{01}^h is actually, as one might guess, the geodesic in M_h between $I^h_{\theta_0}$ and $I^h_{\theta_1}$. Hence

$$\frac{G(I^h_{\theta_0}, I^h_{\theta_1}; M_h)}{L(\gamma_\tau^h; M_h)} = \frac{A_h\|\theta_0 - \theta_1\|}{A_h} = \|\theta_0 - \theta_1\|.$$

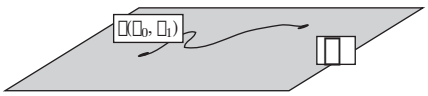

Figure 2: The abstract manifold $\mathcal{M}$

In short, the dependence on h vanishes, and the result is a stable, well-defined (and interpretable!) quantity. The argument given above for translations of the indicator disk is specialized for this particular example. However, in Appendix A we show that the conclusion that renormalization works is valid in a more general context, and that the renormalized limit is finite and defines a sensible metric in a broad range of cases.

3 The Abstract Manifold $\mathcal{M}$

Inspired by the success of renormalization in the special case above, we propose the following shift of viewpoint. We define an abstract manifold $\mathcal{M}$ which consists of a pair (Θ, δ), where Θ is the parameter family underlying M, and δ is a notion of distance on $\Theta \times \Theta$ which is derived from geodesic distance on the smoothed manifolds by the above limiting procedure.

Definition 3.1 *Let $\vartheta(t)$ denote a path in both the abstract manifold $\mathcal{M}$ and in the parameter space Θ. Construct a family M_h of smooth manifolds, where h is the smoothing parameter, converging in L^2 distance to M as $h \to 0$. Let γ^h denote the corresponding path in M_h, and let γ^h_τ be the path in M_h induced from the line segment $\vartheta(\tau)$ in Θ. Then define* **renormalized length** *λ_ϑ in $\mathcal{M}$ by*

$$\lambda(\vartheta; \mathcal{M}) = \lim_{h \to 0} \frac{L(\gamma^h; M_h)}{L(\gamma^h_\tau; M_h)}.$$

How does this work in our running example of translating the disk? Fix landmarks $\tau_0 = (0, 0)$ and $\tau_1 = (1, 0)$; consider for our path the line segment $\vartheta(t) = \theta_0 + t(\theta_1 - \theta_0)$ which runs from θ_0 to θ_1 according to a straight line. We find that

$$\lambda(\vartheta; \mathcal{M}) = \|\theta_1 - \theta_0\|.$$

Definition 3.2 *In referring to* **renormalized geodesic distance** *on $\mathcal{M}$ we intend the follow-*

ing: fix two landmarks τ_0 and τ_1, construct a sequence M_h of smooth manifolds, where h is the smoothing parameter, and which converge in L^2 distance to M as $h \to 0$. Then set

$$\delta(\theta_0, \theta_1; \mathcal{M}) = \min\{\lambda(\vartheta(\cdot); M) : \vartheta(0) = \theta_0, \vartheta(1) = \theta_1\}.$$

Again, in our running example of translating the disk indicator, the length of every line segment in the abstract manifold is the same as its Euclidean length. It follows that the geodesics for δ will just be the Euclidean geodesics. We also propose the following interpretation of δ:

Definition 3.3 *Suppose that, for a constant $c > 0$, the renormalized geodesic distance δ obeys $\delta(\theta_1, \theta_2) = c\|\theta_1 - \theta_2\|_2$, for all $\theta_1, \theta_2 \in \Theta$. Then we say that (Θ, δ) and $(\Theta, \|\cdot\|)$ are in isometry.*

In the rest of this paper, we will study only models of images with edges. We therefore will abandon Definition 2.1, and the associated viewpoint, in favor of this new definition of isometry. No longer do we think of the submanifold M of L^2, instead we think of the abstract manifold $\mathcal{M}$ modeled on Θ, with its own metric. When isometry holds under these definitions, we will say that Continuum Isomap works.

The new definition really matches the original purpose of our study: applications of Isomap to images. An object in a scene is rendered into a digital image by pixelization, and this process provides a limited-resolution image much as $\phi_h \star I_0$ is a limited-resolution image. In some sense applying Isomap to sufficiently fine but limited-resolution imagery gives geodesic distances heuristically of the form $G(\cdot, \cdot; M_h)$. But, as we show in the proof of Theorem 3.1, those distances obey

$$G(\theta_1, \theta_0; M_h) \sim \delta(\theta_1, \theta_0; \mathcal{M}) h^{-\frac{1}{2}}, \qquad h \to 0.$$

Obviously, the factor $h^{-1/2}$ is completely transparent to all operations involved in metric multidimensional scaling, and will not change the reconstruction of pointsets except for a homothetic dilation that can easily be 'standardized away' by fixing the distance between landmark points. This suggests that what really matters for Isomap is the leading coefficient: $\delta(\cdot)$.

There are several things that might be checked about our definition of δ: first, the choice of smoothing is not specified, and the definition might conceivably not be invariant to choice of regularization, so that different regularizations might conceivably lead to different answers; second, this might not define an actual distance; third, the manifold might not be continuous or smooth according to this distance. While there are satisfactory general answers to these questions, we prefer simply to say for now that in specific cases we have studied, these issues do not arise. See also [8].

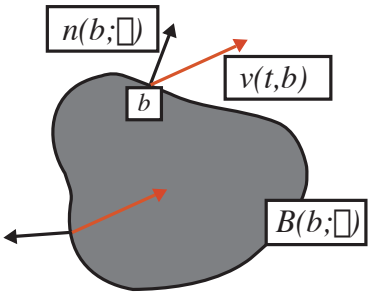

Figure 3: Generic object undergoing a transformation by motion vector $\nu(b,t)$; $n(b;\theta)$ is the boundary normal, $\nu(b,t)$ is the motion vector.

3.1 Calculation of λ, δ

We now give a formula for the length of paths through the image manifold deriving from a family of articulated images. Our formula applies to a range of articulation types. In particular, we will not assume a particular form (such as translation or rotation) for the articulations; merely that they involve smooth transformations of the object in question.

In our model, we consider black-and-white images where the black region B_θ is a kind of 'blob' – a compact set in the plane, with simple closed curve for boundary. The images themselves are indicator functions $I_\theta(\mathbf{x}) = 1_{B_\theta}(\mathbf{x})$, with white represented by 0 and black by 1. The boundary ∂B_θ is a smooth curve.

In defining the regularization procedure, we will use the 2-dimensional standard Gaussian kernel $\phi(x_1, x_2) = \exp\{-(x_1^2 + x_2^2)/2\}$, and, with $\phi_h(\mathbf{x}) = h^{-2}\phi(h^{-1}\mathbf{x})$, define $I_\theta^h = \phi_h \star I_\theta$.

The set B_θ has an outward-pointing normal at each point b of the boundary, labeled $n(b;\theta)$. Each boundary point also undergoes a motion with changing θ. We use the notation $\nu_i(b,\theta)$ for components of the 'motion vector': the rate of change of a specific boundary point with change in the i-th component of the parameter vector θ. This is defined carefully in the Appendix.

We have the following results.

Theorem 3.1 *Let $(\vartheta(t) : t \in [0,1])$ be a smooth curve in parameter space and, using the notation above, set*

$$\nu(b,t) = \sum_i \frac{d\vartheta_i}{dt}\nu_i(b,\vartheta(t)).$$

Supposing that the boundary is smooth and that its motion is smooth, the length of the curve in $\mathcal{M}$ is

$$\lambda(\vartheta) = C_\tau \cdot \int_0^1 \left[\int_{\partial B_\theta} \langle n(b), \nu(b,t)\rangle^2 \, db\right]^{1/2} dt. \tag{3.5}$$

Here C_τ depends on the choice of landmark points, τ_1 and τ_2, used in defining the renormalized length.

Theorem 3.2 *Let $(\vartheta(t) : t \in [0,1])$ be a smooth curve in parameter space and, using the notation above, set*

$$g_{ij}(\theta) = \int_{\partial B_\theta} \langle n(b), \nu_i(b;\theta)\rangle \cdot \langle n(b), \nu_j(b;\theta)\rangle \, db$$

Then g_{ij} defines a Riemannian structure on Θ and the renormalized length of the induced curve in M is the same as the length of the curve in Θ with respect to this Riemannian structure:

$$\lambda(\vartheta) = C_\tau \cdot \int_0^1 \sqrt{\sum_{ij} g_{ij}(\theta) \frac{d\theta_i}{dt} \frac{d\theta_j}{dt}} dt. \tag{3.6}$$

Here C_τ depends on the choice of landmark points used in defining the renormalized length.

We outline the proof for the first of these results in Appendix A; the other is proved similarly; a complete discussion is given in [8]. To see how the results can be applied, consider again our running example of translations of the disk. In this case, boundary points b can be parametrized by the disk center θ and the arclength ω along the perimeter, and

$$b = \theta + (\cos(\omega), \sin(\omega)),$$

while

$$n(b;\theta) = (\cos(\omega), \sin(\omega)),$$

while

$$\nu_i(b;\theta) = \begin{cases} (1,0) & i = 1, \\ (0,1) & i = 2. \end{cases}$$

Again, if we translate from θ_0 to θ_1, the linear path $\vartheta(t) = \theta_0 + t(\theta_1 - \theta_0)$, and so

$$\nu(b,t) = (\theta_1 - \theta_0).$$

Plugging all these formulas into the inner integral in (3.5) gives

$$\int_{\partial B_\theta} \langle n(b), \nu(b,t)\rangle^2 db = \|\theta_1 - \theta_0\|^2 \int_0^{2\pi} \cos(\omega)^2 d\omega = \pi \cdot \|\theta_1 - \theta_0\|^2.$$

Hence we have, as expected:

$$\lambda(\theta) = C_\tau \cdot \|\theta_1 - \theta_0\|.$$

As far as the Riemannian structure goes, we have

$$g_{11}(\theta) = \int_0^{2\pi} \cos(\omega)^2 d\omega = \pi,$$

and similarly $g_{12} = g_{21} = 0$, while $g_{22} = \pi$. Hence the Riemannian structure is simply Euclidean and isometry holds.

3.2 Inferring Euclidean Structure

Below, we will see several examples where g_{ij} is constant and proportional to the identity. This would seem to indicate that the abstract manifold is in fact isometric to a subset of Euclidean space, and that therefore Isomap works. In order to make such an inference from local characteristics to global ones, however, we need an extra element. In order to be globally Euclidean, the local structure must be Euclidean and the space Θ must be convex. If it is not convex, then shortest paths in Θ will not be dictated by the infinitesimal metric structure alone, but will in some cases be dictated by the nonconvexity of the space, e.g. to avoid holes. Therefore, we will sometimes emphasize the issue of convexity.

In some cases, one can demonstrate that isometry holds without computing either δ or g. In fact, all that is necessary is that Θ be convex, and that, for every pair of points $\theta_i \in \Theta$, the linear path in parameter space $\vartheta(t) = \theta_0 + t(\theta_1 - \theta_0)$ has a length proportional to the Euclidean distance $\|\theta_1 - \theta_0\|$, where the constant of proportionality does not depend on θ_i.

4 Simple Examples

We now give a few simple examples of the preceding theory.

4.1 Translating a 4-fold symmetric object

We now generalize the problem of translating a disk, considering a broader class of symmetric objects.

Definition 4.1 *An object B_0 has 4-fold symmetry if it is invariant under rotations about the origin by* 90*,* 180*, and* 270 *degrees.*

This class of objects includes ℓ^p balls for $0 < p < \infty$:

$$\ell^p = \{\mathbf{x} : |x_1|^p + |x_2|^p \leq 1\},$$

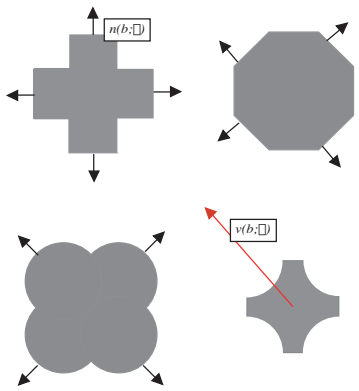

Figure 4: Translation of a 4-fold symmetric object

although for the cases $p =\in \{1, \infty\}$ the boundaries are not smooth.

Theorem 4.1 *For translations of 4-fold symmetric figures with smooth boundary, isometry holds; the renormalized geodesic distance is proportional to the Euclidean distance in parameter space.*

As ℓ^p balls for $p = 2$ are just disks, this contains our earlier result about translating disks as a special case. However, this result is much more general than just to ℓ^p balls; it works for a wide variety of nonconvex figures; see Figure 4.

Proof. Consider the integrand of (3.5):

$$\int_{\partial B_\theta} \langle n(b), \nu(b,t)\rangle^2 db$$

Now for a translation family, and a path $\vartheta(t) = \theta_0 + t(\theta_1 - \theta_0)$, we have that $\nu(b,t) = \theta_1 - \theta_0$; this is a constant, v, say, independent of t and b. Hence the integral becomes:

$$\int_{\partial B_\theta} \langle n(b), v\rangle^2 \, db$$

Now for a 4-fold symmetric object, let Q_θ denote the northeast boundary (i.e. the intersection of ∂B_θ with the positive orthant). Hence, we may write (taking $b_0 = b$)

$$\int_{\partial B_\theta} \langle n(b), v\rangle^2 db = \int_{Q_\theta} \sum_{k=0}^{3} \langle n(b_i), v\rangle^2 db$$

Now for every $b \in Q_\theta$,

- there are three corresponding boundary points b_1, b_2, b_3 obtained by rotations through 90,

180, and 270 degrees;

- we have antisymmetric pairs of normals at b, b_2 and at b_1, b_3, in the sense that the respective normals $n(\cdot)$ point in opposite directions at each member of the pair; and
- we have orthogonal pairs of normals at b, b_1, and at b_2, b_3 in the sense each pair of normals $n(\cdot)$ is mutually orthogonal.

Now letting $n^{\perp}(b)$ be perpendicular to $n(b)$, we have

$$\sum_{k=0}^{3} \langle n(b_i), v\rangle^2 = 2(\langle n(b), v\rangle^2 + \langle n^{\perp}(b), v\rangle^2) = 2\|v\|^2.$$

Hence,

$$\int_{\partial B_\theta} \langle n(b), v\rangle^2 db = \frac{\|v\|^2}{2} \int_{\partial B_\theta} db = \frac{\|v\|^2}{2} L(\partial B_0)$$

In short, recalling the definition of $v = \theta_1 - \theta_0$:

$$\lambda(\vartheta) = \sqrt{L(\partial B_0)/2} \cdot \|\theta_1 - \theta_0\|.$$

This implies the proportionality of geodesic and Euclidean distance. □

Vin de Silva explained to us that a more general result is true, in which 4-fold symmetry is replaced by k-fold symmetry for any $k \geq 3$. De Silva's argument starts by noting that the infinitesimal structure of the metric δ is quadratic, and so has ellipsoidal level sets. At the same time, k-fold symmetry of B_0 would impose k-fold symmetry of that ellipsoid, which, for $k \geq 3$, forces it to be circular, and hence makes the infinitesimal structure conformal. Isometry follows by noting that the scale of the problem is translation-invariant.

4.2 Pivoting an Object

Consider an object B_0 with smooth boundary ∂B_0, and suppose that $0 \in \partial B_0$. Now consider a family of articulations that pivot the object around zero. More formally, let R_θ denote the rotation matrix

$$R_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix},$$

let $I_0(\mathbf{x}) = 1_{B_0}(\mathbf{x})$, and let

$$I_\theta = I_0(R_\theta \mathbf{x}).$$

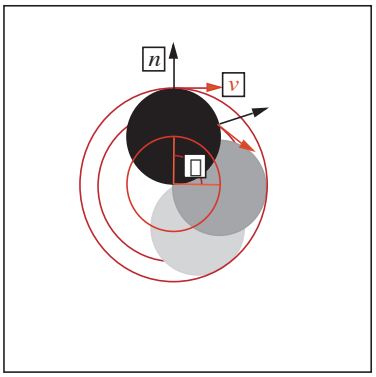

Figure 5: Pivoting a disk around a boundary point

Theorem 4.2 *In the Pivot model, isometry holds; for* $|\theta_0 - \theta_1| < \pi$ *renormalized geodesic distance is proportional to the Euclidean distance in parameter space.*

Proof: One has merely to notice that $g_{11}(\theta)$ is independent of θ. This is easy to see by observing that the integrand in g_{11},

$$\langle n(b), \nu_1(b;\theta)\rangle^2$$

is independent of θ. Indeed, for any value of θ, $n(b;\theta)$ points normal to ∂B, while $\nu_1(b;\theta)$ points normal to the line segment joining 0 to b. The angle between these two vectors is the same independent of θ (See Figure 5).

4.3 Extensions to Piecewise Smooth Boundaries

The results above can be generalized to the case of piecewise smooth boundary curves.

On the one hand, as a limiting case of Theorem 4.1, consider translations of an ℓ^p ball for $p = \infty$, where B_0 is the unit square in the plane. Extrapolating the above result from $p < \infty$ to $p = \infty$ suggests that it might also be true that isometry holds for translations of the square, and indeed this is the case. To verify that, we need to know that (3.5) still works if ∂B_θ is merely piecewise smooth. The demonstration of this appears in [8]. It is clear that the same extension applies not only to squares, but to $4k$-gons, for any k (squares are the case $k = 1$. It is also clear that the same extension applies to the case of objects with piecewise smooth boundaries and 4-fold symmetry. In all cases, isometry holds.

On the other hand, as a limiting case of Theorem 4.2, consider a 'pie slice' B_0 outlined by two radial line segments, $\{(r\cos(+\omega), r\sin(+\omega)) : 0 \le r < R\}$ and $\{(r\cos(-\omega), r\sin(-\omega)) :$

$0 \leq r < R\}$; and an arc connecting them, $\{(R\cos(\omega'), R\sin\omega') : -\omega \leq \omega' \leq \omega\}$.

Articulate the structure by a pivoting it according to R_θ as in Section 4.2 above. Hence, with $I_0 = 1_{B_0}$, set $I_\theta(\mathbf{x}) = I_0(R_\theta \mathbf{x})$.

This is a limiting case of Theorem 4.2 because the boundary is piecewise smooth rather than smooth. However, because (3.5) extends to the case of objects with piecewise smooth boundaries, the conclusion of Theorem 4.2 carries over to this case as well.

5 Empirical Results

In this section, we compare our theoretical results showing isometry in a continuum setting against empirical results applying Isomap to families of digital images.

5.1 Rigid translation of a single disk

Consider our standing example where the image $I_0(x) = 1_{\{|x|\leq 1\}}$ is the indicator of a unit disk, and our articulation translates the center of the disk by a vector θ in $\mathbb{R}^2$. As indicated above, isometry holds in the continuum setting. We expect that an empirical trial with a finite database of digital imagery should show essentially 'perfect' recovery of the original parameter space.

Our test uses the publicly-available Isomap code ([1]), running with the nearest-neighbor parameter $k = 7$, and a set of 100 black-on-white images at 64-by-64 pixels. The translation of the prototype was performed by randomly selecting x, y-coordinates of the disk centers, restricting the centers so that the disk stayed inside the borders of the image. The results of the unit disk experiment are seen in Figure 6. The actual Isomap 2-dimensional embedding was optimally rotated and scaled using a Procrustes rotation [7] to match it to the original parameter space for graphical presentation. Isomap almost perfectly recovers the original spacing and relative positioning of the disk centers. In fact, we note that the only noticeable deviations tend to occur where the sampling was particularly sparse, such as in the upper left corner of the parameter domain.

5.2 Pivoting of a Single Disk

Consider pivoting a rotating a unit disk around a fixed boundary point, which for convenience is chosen as the origin of our coordinate system. The position of the center of the disk provides a natural parametrization; as it lies on the unit circle, it can be measured in radians θ. The

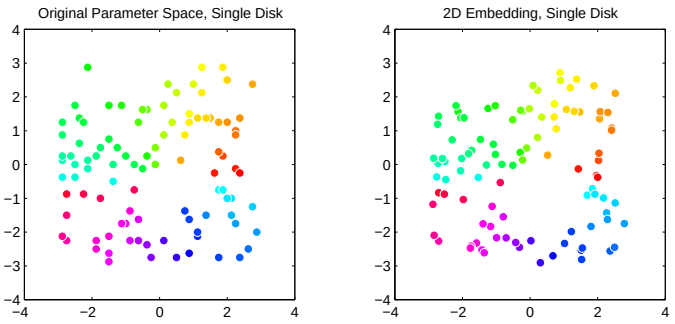

Figure 6: Left panel: Original Disk Centers, colored by polar angle. Right panel: 2-dimensional embedding recovered by Isomap. The points in the right panel have been colored by original polar angle (optimally rotated and scaled). Similarly-colored points are typically in similar positions in the two panels.

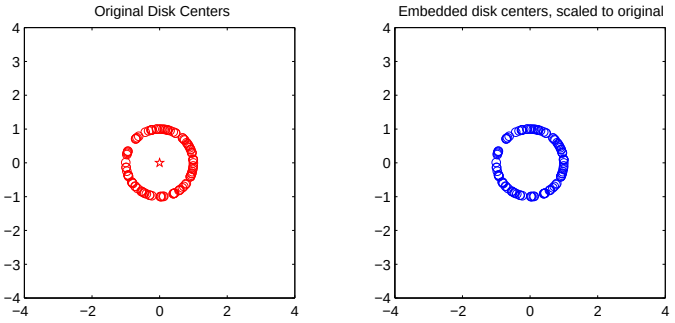

Figure 7: Right Panel: Original disk centers. Left Panel: 2-dimensional embedding for single disk pivots

renormalized length of a line segment $[\theta_0, \theta_1]$ is given by (3.5) after some calculation as

$$\lambda(\vartheta) \propto \pi^{\frac{1}{2}} \times |d\theta|$$

where the articulation is a pivoting through an angle $d\theta = |\theta_0 - \theta_1|$. As in the previous experiment, we create a database of 100 images of a black-on-white disk at 64-by-64 pixels, and articulate the disk under pivoting through a randomly selected angle θ.

The resulting experiment, optimally rotated and scaled, shows a satisfactory match to the original parameter space in Figure 7. (We plot the disk centers in two-dimensions for comparison to the image field).

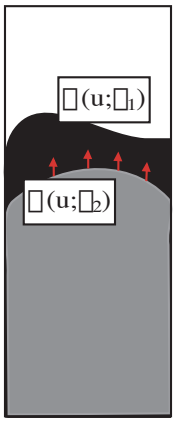

Figure 8: Horizon Articulation Model

6 Variations

In this section, we return to the continuum setting and give more complex examples of articulation/object models exhibiting isometry.

6.1 Horizon Articulation

We consider first a variation in both the image index set and the articulation model. Now the image $I(u, v)$ will have domain $0 \leq u \leq 1$ and $-\infty < v < \infty$, forming a vertical 'strip', and the images of interest to us will be black and white images with the boundary between the two colors being a horizon $v = \psi(u)$. Hence

$$I(u, v; \psi) = 1_{\{v \leq \psi(u)\}}.$$

We are interested in families I_θ of articulated images where the horizon is generated by a linear combination of basis elements ψ_j (an example of an articulation appears in Figure 8)

$$\psi(u; \theta) = \sum_i \theta_i \psi_i(u).$$

Hence θ is a vector of expansion coefficients governing the shape of the horizon. We assume the basis functions are orthogonal for $L^2[0, 1]$:

$$\int_0^1 \psi_j(u)\psi_{j'}(u)du = 1_{\{j=j'\}};$$

examples include orthonormal wavelets, and orthonormal sinusoids. We will assume that the

basis functions are smooth: at least C^2.

In defining the metric δ for this setting, we need to define image regularization. The natural choice in this case is to smooth 'vertically'. So let φ denote the 1-dimensional standard normal density, and the smoothing operation

$$(\varphi_h \star_v f)(u,v) = \int_{-\infty}^{\infty} \varphi_h(w) f(u, v-w) dw;$$

define then the smoothed image I^h by

$$I^h_\theta = \varphi_h \star_v I_\theta.$$

For fixed $h > 0$, the collection of all I^h_θ for all $\theta \in \Theta$ defines a smooth image manifold M_h, and so we can define renormalized length in a fashion similar to (3.5). In this setting, the normal $n(b)$ is no longer relevant; we define the motion vector

$$v(t,u) = \sum_j \frac{d\theta_j}{dt} \psi_j(u);$$

and the formula for renormalized length is

$$\lambda(\vartheta) = C_\tau \cdot \int_0^1 \left[\int_0^1 v(t,u)^2 du \right]^{1/2} dt. \tag{6.7}$$

When we do this, we find that, if we select a linear path in parameter space,

$$\vartheta(t) = \theta_0 + t(\theta_1 - \theta_0)$$

then

$$\lambda(\vartheta) = C_\tau \cdot \|\theta_1 - \theta_0\|.$$

Since the length of every line segment is proportional to the Euclidean distance, it follows that δ itself is proportional to Euclidean distance.

Also, the appropriate formula for the Riemannian metric is simply

$$g_{ij}(\theta) = C_\tau^2 \cdot \int_0^1 \phi_i(u)\phi_j(u) du = C_\tau^2 \cdot 1_{\{i=j\}}.$$

which shows the same thing another way.

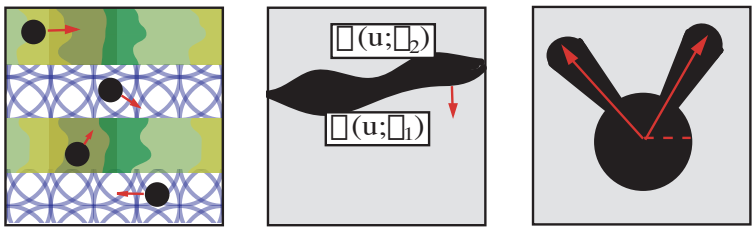

Figure 9: Successful Composite Articulations: Multiple Ordered Disks, Horizon models, and Bunny Ears

Theorem 6.1 *In the Horizon Articulation model, if the parameter space* Θ *is convex, isometry holds.*

An interesting feature here is that there is no limit to the dimensionality of θ; in particular, we can consider quite complex articulations of a horizon; in each case, the natural parametrization will be recovered up to a rigid motion of parameter space.

6.2 Bunny Ears and Hands

We now consider multiple-object articulations. Restrictions on the joint behavior of the objects will be essential in order to have isometry. For detailed discussion of the necessary conditions and proofs of the results, see [8].

Consider a two-object 'Bunny Ears' composed of two wedges on a disk: Let $\theta = (\theta^1, \theta^2)$ denote a pair of angles, and let B_θ be the unit disk with two 'ears' attached, with the inclination of the 'ears' controlled by the components of θ. Each ear is an annular wedge with angular opening 2ω. Such a basic wedge is defined by boundary segments $\{(r\cos(\theta+\omega), r\sin(\theta+\omega)) : 1 \le r < 2\}$ and $\{(r\cos(\theta-\omega), r\sin(\theta-\omega)) : 1 \le r < 2\}$; and two arcs connecting them, along $r = 1$ and $r = 2$. The first ear is a wedge centered at $\theta = \theta^1$; the second ear is a wedge centered at $\theta = \theta^2$. The parameter space Θ is chosen so the ears don't overlap:

$$\Theta = \{(\theta^1, \theta^2) : 2\omega < \theta^1 - \theta^2 < 2\pi - 2\omega\}.$$

We can extend this to the *Hands* model, in which the unit disk has 5 'fingers' attached, and each 'finger' is again an annular wedge at specific angle θ^i, $i = 1, \ldots, 5$.

$$\Theta = \{(\theta^1, \theta^2, \ldots, \theta^5) : 2\omega < \theta^i - \theta^{i+1} < 2\pi - 2\omega\}.$$

As long as the condition that the mobile wedges cannot overlap is enforced, the demonstration

can be given for any number of 'fingers'.

Theorem 6.2 *For the 'Bunny Ears' and 'Hand' Models, with the indicated convex parameter spaces, isometry holds.*

6.3 Translating Several Ordered Nonoverlapping Disks

Consider now continuum images containing n disks, each of radius one, with centers θ_i, $i = 1, \ldots, n$. With I_0 the indicator of the unit disk, and $\theta = (\theta_1, \ldots, \theta_n)$, put

$$I_\theta(\mathbf{x}) = \sum_{I=1}^{n} I_0(\mathbf{x} - \theta_i).$$

In this section we constrain

$$\|\theta_i - \theta_j\| > 1$$

so that no two terms in the sum overlap, and the image remains of the 'black-and-white' type in previous sections. We will go further than merely constrain the disks to be nonoverlapping; we do so in several ways, each defining a different parameter space Θ.

- *Single File.* Let Θ_1 denote the set of θ where the disk centers lie along the x-axis, so that

  $$\theta_{i,2} = 0 \qquad i = 1, \ldots, n;$$

  (here $\theta_{i,j}$ denotes the j-coordinate of θ_i); moreover the centers are ordered along the x-axis:

  $$\theta_{1,1} < \theta_{2,1} < \cdots < \theta_{n,1}.$$

- *Separated Columns.* Let $\Theta_2 = \Theta_2((a_i))$ denote the set of θ arising where the disk centers can be distributed in the 2-dimensional plane relatively broadly, but are constrained to lie in zones defined by regions of the x-axis. We fix cutpoints a_0, a_1, ..., a_n, with $a_i > a_{i-1}+2$ and demand that

  $$a_{i-1} + 1 < \theta_{i,1} < a_i - 1 \qquad i = 1, \ldots, n.$$

- *Northeast/Southwest ordering.* We let Θ_3 denote the parameter space where the disks are arranged so that the i-th one has all earlier disks lying to the southwest and all later disks lying in the northeast quadrant.

Each such constraint corresponds to a *convex* family Θ. Now we consider independent translations of the family of disks that maintain the assumed constraint, i.e. translations that preserve membership in Θ.

In this setting, we can adapt arguments from the case of a single disk to see that, if $\theta_{(0)} = (\theta_0^1, \ldots, \theta_0^n)$ is a collection of disk centers and $\theta_{(1)} = (\theta_1^1, \ldots, \theta_1^n)$ is another collection of disk centers, then for $\vartheta(t) = \theta_{(0)} + t(\theta_{(1)} - \theta_{(0)})$, we have the renormalized length:

$$\lambda(\theta; \mathcal{M}) = C_\tau \cdot \sqrt{\sum_{i=1}^{n} \|\theta_1^i - \theta_0^i\|^2}$$

and a Riemannian structure which is in fact Euclidean:

$$g(\theta) = C_\tau \cdot Id, \qquad \forall \theta \in \Theta.$$

Thus, in this setting, the shortest distance between pairs of images indeed goes by morphing the centers along line segments.

Theorem 6.3 *Isometry holds for the Ordered Nonoverlapping Disks model with any of the parameter spaces Θ_1, Θ_2, or Θ_3.*

The extension from Disks to 4-fold symmetry seems likely to hold as well.

6.4 Two Non-Crossing Horizons

Now consider a region which is white everywhere except between two horizons which are constrained never to cross. For example, suppose that, as in Figure 9, both horizons are touching at the endpoints $u = 0, 1$, and that elsewhere $\psi_1(u) < \psi_2(u)$. Suppose that the two horizons are each parametrized as in Section 5.1, with respective parameter vectors θ_i, and that, with $\theta = (\theta_1, \theta_2)$, we have

$$I_\theta = 1_{\{\psi(u;\theta_1) \le v \le \psi(u;\theta_2)\}}.$$

We let Θ be the subset of θ vectors where

$$\psi(u; \theta_1) \le \psi(u; \theta_2) \qquad u \in [0, 1],$$

so that the definition makes sense. Note that, for each fixed u, $\theta \mapsto \psi(u; \theta)$ is linear; therefore, Θ is a convex parameter space.

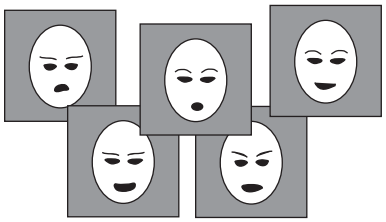

Figure 10: Examples of cartoon faces generated by horizon articulations

It is not hard to see that if $\theta_{(0)} = (\theta_0^1, \theta_0^2)$ is a collection of Horizon parameters in Θ and $\theta_{(1)} = (\theta_1^1, \theta_1^2)$ is another collection of horizon parameters in Θ, then for $\vartheta(t) = \theta_{(0)} + t(\theta_{(1)} - \theta_{(0)})$ a linear path we have the arclength

$$\lambda(\vartheta; \mathcal{M}) = C_\tau \cdot \sqrt{\sum_{i=1}^{2} \|\theta_1^i - \theta_0^i\|^2}$$

and a Riemannian structure which is in fact Euclidean:

$$g(\theta) = C_\tau^2 \cdot Id, \qquad \forall \theta \in \Theta.$$

Thus, again in this setting, the shortest path between images defined by two horizons indeed goes by following a linear path in parameter space.

Theorem 6.4 *Isometry holds for the Two Non-Crossing Horizons model.*

6.5 Cartoon Faces

Now consider a very simple model of a cartoon face undergoing various articulations.

A fixed oval region in the plane is called the 'head', and five regions inside it, called 'brows', 'eyes' and 'mouth' are defined. Each of the brows, eyes and mouth is a black region defined by two articulating horizons. The eyes and mouth have horizons which, as in the previous section, are joined at the ends, creating an almond shape. The brows have two horizons ending in vertical line segments and having fixed widths. Examples of these faces appear in Figure 10. The parameter space under this model is convex for the same reasons as the parameter space in Section 6.4.

The underlying parameter vector $\theta = (\theta^1, ..., \theta^8)$; θ^1 and θ^2, give the upper and lower horizon of the mouth (i.e. the lips), θ^3 and θ^4 give the upper and lower horizon of the left eye, (i.e. the lids), etc. The parameter space Θ is constrained so that the upper and lower boundaries of an

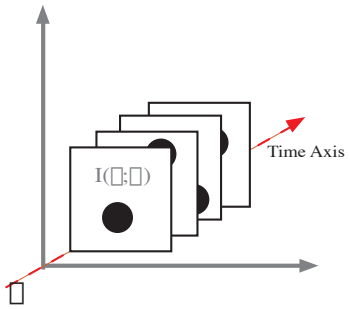

Figure 11: Movie model: sequence of articulated disks in time

object such as the right eye never cross, and so that the upper boundary of the right eye never crosses the lower boundary of the right brow, etc.

We evaluate closeness of the images using $L^2(d\mathbf{x})$ as usual. Applying the same smoothing recipe as in the single horizon case, we get a renormalized distance.

It is not hard to see that if $\theta_{(0)} = (\theta_0^1, \ldots, \theta_0^8)$ is a collection of Horizon parameters in Θ and $\theta_{(1)} = (\theta_1^1, \ldots, \theta_1^8)$ is another collection of horizon parameters in Θ, then for a linear path $\vartheta(t) = \theta_{(0)} + t(\theta_{(1)} - \theta_{(0)})$ we have the arclength

$$\lambda(\vartheta; \mathcal{M}) = C_\tau \cdot \sqrt{\sum_{i=1}^{8} \|\theta_1^i - \theta_0^i\|^2}$$

and a Riemannian structure which is in fact Euclidean:

$$g_{ij}(\theta) = C_\tau^2 \cdot Id, \qquad \forall \quad \theta \in \Theta.$$

Thus, again in this setting, the shortest path between images defined by two horizons indeed goes by following a linear path in parameter space.

Theorem 6.5 *Isometry holds for the Cartoon Face Model.*

7 Movies

In this section only, we consider the case of 'movies', which are objects with an additional time index: $I(\mathbf{x}, \alpha)$ where $\mathbf{x} \in \mathbb{R}^2$ and the time index $\alpha \in [0, 1]$. Think of a (continuous) collection of 2-D images $I(\cdot, \alpha)$; instead of looking at the parameter space for one movie, we consider the joint parameter space of an entire set of movies.

We consider two kinds of articulations under the movie model. For much more ambitious

examples, see the technical report [9].

7.1 Motions of a Disk

Let I_0 be a 2-dimensional image – the indicator function of a disk – and consider a movie showing motion of the disk. Let in fact $\chi(\alpha)$ be the center of the disk as a function of time and put

$$I(x, \alpha) = I_0(x - \chi(\alpha)).$$

For a simple example, consider a movie which shows linear motion of the disk. Then

$$\chi(\alpha) = \nu \cdot \alpha.$$

The center of the disk is at $(0, 0)$ at the movie's beginning, at time $\alpha = 0$, and it is at ν at the movie's end, at time $\alpha = 1$.

For a more interesting class of motions, we consider the parametric family:

$$\chi(\alpha) = \chi(\alpha; \theta) = \sum_j \theta_j \psi_j(\alpha)$$

where ψ_j are vector-valued and orthogonal in $L^2[0, 1]$.

Define now the family of articulated images $I_\theta(\mathbf{x}) = I_0(\mathbf{x} - \chi(\alpha; \theta))$, and for regularization, do 2-d smoothing within each frame:

$$I_\theta^h = \phi_h \star_{(x,y)} I_\theta;$$

here ϕ_h is a 2-d smoothing kernel in (x, y) only (not in α). For a smooth path $\vartheta(\cdot)$ in Θ we have the following formula for arclength:

$$\lambda(\vartheta) = C_\tau \int_0^1 \sqrt{\int \langle \nu, n \rangle^2 db d\alpha}\, dt$$

where b runs along the perimeter of the disk, $\nu = \nu(b, \alpha; t)$ is the motion vector of the boundary element b at time α in the movie and at parameter t in the path, and n is the (x, y)-normal to the surface of the disk at (b, α). Then we notice that, just as in the case of translating a 2-d image,

$$\int \langle \nu, n \rangle^2 db = C \cdot \| \frac{d\chi}{d\alpha} \|_2^2$$

We conclude that

$$\delta(\theta_0, \theta_1) = C_\tau \cdot \|\theta_1 - \theta_0\|_2.$$

Theorem 7.1 *Isometry holds for movies of a translating disk.*

For similar reasons, Continuum Isomap works for movies of several translating disks which are ordered and nonoverlapping.

7.2 Gestures of a Hand

Consider the Hands model of the previous section, only relabel the parameter θ of that model by ζ. Then ζ has several components, controlling the position of the various fingers. Now consider a movie showing motion of the hand components. Let in fact $\zeta(\alpha)$ be the parameters of the hand as a function of time and put

$$I(\mathbf{x}, \alpha) = I_{\zeta(\alpha)}(\mathbf{x}).$$

For a simple example, consider a movie which shows uniform rotation of the fingers around a common center. Then

$$\zeta(\alpha) = \zeta_0 + \alpha(\zeta_1 - \zeta_0).$$

The hand is in configuration ζ_0 at the movie's beginning, at time $\alpha = 0$, and it is at ζ_1 at the movie's end, at time $\alpha = 1$.

For a more interesting class of motions, we consider the parametric family:

$$\zeta(\alpha) = \zeta(\alpha; \theta) = \sum_j \theta_j \zeta_j(\alpha)$$

where ζ_j are vector-valued and orthogonal in $L^2[0, 1]$.

Define now the family of articulated images $I_\theta(\mathbf{x}, \alpha) = I_{\zeta(\alpha)}(\mathbf{x})$. and for regularization, do 1-D smoothing angularly:

$$I_\theta^h = \phi_h \star_\omega I_\theta$$

here ϕ_h is a 1-D smoothing kernel acting convolutionally in r = constant. For a smooth path $\vartheta(\cdot)$ in Θ we have the following formula for arclength:

$$\lambda(\vartheta) = \int_0^1 \sqrt{\int \langle \nu, n \rangle^2 db \, d\alpha} \, dt$$

where b runs across the perimeter of the hand within the two-dimensional image, $\nu(b, \alpha; t)$ is the

motion vector of the boundary element at time α in the movie and at parameter t in the path, and n is the (x, y)-normal to the surface of the disk at (b, α). Then we notice that, just as in the case of translating a two-dimensional image,

$$\int \langle \nu, n \rangle^2 db = C \cdot \| \frac{d\zeta}{d\alpha} \|_2^2.$$

We conclude that

$$\delta(\theta_0, \theta_1) = C \cdot \|\theta_1 - \theta_0\|_2.$$

Theorem 7.2 *Isometry holds for movies of Hand Gestures.*

8 Failures

The conditions established for Continuum Isomap to 'work' can fail in two different ways. First, the parameter space can be inherently non-convex. Second, the manifold can be essentially curved.

8.1 Non-Convexity: Two Disks with Exclusion

Consider images of two black disks on a white background within a square image as in Figure 12. Suppose the disks are forbidden to occupy the same space at the same time. Partitioning the parameter θ as $(\theta^(1), \theta^(2))$, with components denoting the positions of the image subcomponents, the disks cannot overlap. At any given position of the first disk, the parameter space of the second disk has a hole representing the parameter values leading to overlap with the first disk. As a result, the parameter space is not convex. For further examples of non-convexity, see [9].

In this case, the Riemannian structure $g_{ij}(\theta)$ is proportional to the identity everywhere, but the domain is not convex. We have local isometry but not isometry. As it turns out, local isometry is sufficient to allow reconstruction of the parameter space, using one of the methods described in [5].

8.2 Non-Flatness: Rectangles

We now turn to examples where the abstract manifold is not flat. Consider the 'world' of rectangles of various volumes and aspect ratios and common centers. With α_1 and α_2 the halflengths in the axial directions, and $\theta = (\alpha_1, \alpha_2)$, the image I_θ is the the indicator of the region defined by $|x| \leq \alpha_1$, $|y| \leq \alpha_2$. Let Θ be the rectangle $\alpha_i \in (1/2, 2)$.

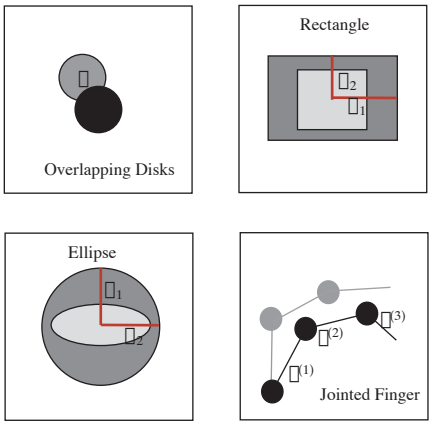

Figure 12: Failures: Rectangle Morphing, Ellipse Morphing, Race Track, Jointed FInger

Now consider a smooth path in Θ. Because at each point of ∂B_θ the normal vector is either coincident with the motion vector or perpendicular to it, and because articulation in α_1 alone has an impact corresponding to the vertical profile of the rectangle,

$$g_{11} = C_\tau \times \Big[\int_{x\in[-\alpha_2,\alpha_2]} dx\Big]$$

In short,

$$\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} C_\tau\alpha_2 & 0 \\ 0 & C_\tau\alpha_1 \end{bmatrix}$$

As the Riemannian structure (g_{ij}) depends in an essential way on θ, it is not constant, and the abstract manifold $\mathcal{M}$ is not flat. Hence, we do not have isometry between the apparently natural parameter space (Θ, δ) and Euclidean space.

Other examples of this type, e.g. images of ellipses with various aspect ratios and volumes, are discussed in [9]. Many three-dimensional perspective transformations result in inherent curvature of the abstract manifold.

8.3 Non-flatness: Fingers with Multiple Joints

Imagine a model as in the fourth panel of Figure 12 where we define a single 'skeletal finger' composed of three connected line segments, thickened slightly into 'rods'. In this case, we assume that the joints of the finger are negligible compared to the segment lengths (all segments are

of length L). Parametrize each of the segments by an angular position: θ_i, $i = 1, 2, 3$, where $0 \leq \theta_i \leq \pi/2$. The entire skeleton of the finger is then, for $l \in [0, L]$:

$$\begin{aligned}
S_1 =& (l\cos(\theta_1), l\sin(\theta_1)) \\
S_2 =& (L\cos(\theta_1) + l\cos(\theta_2), L\sin(\theta_1) + l\sin(\theta_2)) \\
S_3 =& (L\cos(\theta_1) + L\cos(\theta_2) + l\cos(\theta_3), L\sin(\theta_1) + L\sin(\theta_2) + l\sin(\theta_3)))
\end{aligned}$$

As a result, the normal and motion vectors under a three-parameter articulation, $(\theta_1, \theta_2, \theta_3)$, depend on the values of the other variables. To achieve the minimal distance between two positions of θ_1 requires that the other parameters θ_i, $i > 1$ be in optimal position, and the definition of an 'optimal' position depends on the current value of θ_1. The result is curvature.

9 Conclusion

We have studied manifolds of model images defined over the continuum plane, finding that for a number of interesting articulation families, the resulting abstract image manifolds are isometric to the underlying 'natural' parameter space. This gives a theoretical framework explaining why the Isomap algorithm can be expected to work in the case of image manifolds.

Although our analysis abstracts away several possible confounding factors such as pixelization, the comparison of theoretical predictions with empirical tests of Isomap suggests that the theory is accurate in predicting actual empirical results. In addition, our theoretical perspective allows consideration of variations of the image case (movies and very high-dimensional articulations, for example) that would not be easy to explore by empirical techniques.

A Proof of Theorem 3.1

Notation: Let $I_t \equiv I_{\vartheta(t)}$, where I_θ is a function of $\mathbf{x} \in \mathbb{R}^2$. Let the boundary of the set B be denoted by ∂B; points in the boundary are $b \in \partial B$. We assume that B and ∂B have the same topology throughout Θ, and so, for a fixed θ_0 there exists a parametrizing map $\alpha : \partial B_{\theta_0} \times \Theta \to \mathbb{R}^2$ using ∂B_{θ_0} as a parameter space for each boundary point and each $\theta \subset \Theta$. We also assume that the radius of curvature of ∂B is uniformly bounded away from zero both over every $b \in \partial B_\theta$ and over $\theta \subset \Theta$. It follows from this that for fixed θ_0 there is a tubular neighborhood of ∂B_{θ_0} with coordinates (b, u), where any $\mathbf{x}$ in that neighborhood has for its b-coordinate the closest point in ∂B_{θ_0} whose normal $n(b)$ is parallel to $x - b$, and for its u-coordinate the signed distance to

∂B_{θ_0}. Fix now θ_0 and consider the coordinate system associated to ∂B_{θ_0}. The motion vectors $\nu_i(b,\theta)$ referred to verbally in Section 3 are then defined by

$$\nu_i(b,\theta) = \frac{\partial}{\partial\theta_i}\alpha(b,\theta),$$

and, if θ evolves along a path $\vartheta(t)$, the boundary ∂B_θ flows with time according to

$$\nu(b,t) = \sum_i \frac{d\theta_i}{dt}\nu_i(b,\vartheta(t)).$$

We can also parametrize the boundary ∂B_θ, for all θ sufficiently close to θ_0, in terms of a function $\beta(b,\theta)$ giving the u-coordinate of the boundary point in ∂B_θ corresponding to a given b in ∂B_{θ_0}. We will make the quantitative assumption that β is a C^2 function of both θ and b jointly. The proof is a series of observations recorded in the form of Lemmas.

Lemma A.1 *If ψ is $C^\infty(\mathbb{R}^2)$ and of rapid decay at ∞,*

$$\frac{d}{dt}\langle I_t,\psi\rangle\big|_{t=0} = \int_{\partial B}\sigma(b)\,\psi(b)\,db$$

where $\sigma(b) \equiv \langle\nu(b,0), n(b)\rangle$.

Proof. The Tubular Neighborhood Theorem [10] allows us to write, for small enough $\epsilon > 0$

$$\langle I_\epsilon - I_0,\psi\rangle = \int_{\partial B}\int_0^{\beta(b,\theta_\epsilon)}\psi(b+un(b))J(u,b)\,du\,db$$

where $J(u,b)$ is the Jacobian of the cartesian-to-tubular coordinates map (and which is $1+o(1)$ as $u\to 0$), and where we interpret an integral with reversed limits according to $\int_0^{-|a|} \equiv -\int_{-|a|}^0$ as usual. Now

$$\beta(b,\theta_\epsilon) = \epsilon\sigma(b) + O(\epsilon^2)$$

so

$$\int_0^{\beta(b,\theta_\epsilon)}\psi(b+un(b))J(u,b)\,du \;\sim\; \epsilon\psi(b)\sigma(b), \quad \epsilon\to 0.$$

Dividing by ϵ and letting $\epsilon\to 0$ gives the result. □

Lemma A.2

$$\frac{d}{dt}[\phi_h \star I_t(x)]_{t=0} = \int_{\partial B}\phi_h(x-b)\sigma(b)\,db \tag{1.8}$$

Proof. Define the function ψ_{x_0} according to $\psi_{x_0}(y) = \phi_h(x_0 - y)$. Self-adjointness of

convolution allows us to write

$$\begin{aligned}\frac{1}{\epsilon}(\phi_h \star (I_\epsilon - I_0))(x_0) &= \frac{1}{\epsilon}\int (I_\epsilon - I_0)(y)\psi_{x_0}(y)\,dy \\ &\to \frac{d}{dt}\langle I_t, \psi_{x_0}\rangle\big|_{t=0}, \qquad \epsilon \to 0, \\ &= \int_{\partial B} \psi_{x_0}(b)\sigma(b)\,db\end{aligned}$$

where the last step follows from Lemma (A.1). (1.8) follows. □

Lemma A.3 *Let* $\psi_h(x) \equiv \phi_{\sqrt{2}h}(x)$. *Then*

$$\int (\frac{d}{dt}(\phi_h \star I_t))^2\,dx = \int_{\partial B}\int_{\partial B} \sigma(b)\sigma(b')\psi_h(b-b')\,db\,db'.$$

Proof.

$$\begin{aligned}\int (\frac{d}{dt}(\phi_h \star I_t))^2\,dx &= \int_{\mathbb{R}^2} \Big(\int_{\partial B} \phi_h(x-b)\sigma(b)\,db\Big)\Big(\int_{\partial B} \phi_h(x-b')\sigma(b')\,db'\Big)\,dx \\ &= \int_{\partial B}\int_{\partial B} \sigma(b)\sigma(b')\Big(\int_{\mathbb{R}^2} \phi_h(x-b)\phi_h(x-b')\,dx\Big)\,db\,db' \\ &= \int_{\partial B}\int_{\partial B} \sigma(b)\sigma(b')\psi_h(b-b')\,db\,db'\end{aligned}$$

where ψ_h denotes the cross-correlation $\phi_h \otimes \phi_h$. Because ϕ_h is Gaussian, its cross-correlation with itself, ψ_h, is also Gaussian, with a scale parameter larger by a factor of $\sqrt{2}$. □

Now combine the above lemmas. With ψ an appropriately-scaled isotropic Gaussian,

$$\psi_h(b-b') = \psi\big(\frac{b-b'}{h}\big)\frac{1}{h^2}.$$

Note that near b, ∂B is approximated by $L = \{b + t\dot{b} : t \in \mathbb{R}\}$ where $\dot{b}$ is the unit tangent vector to ∂B at b. Therefore, if $b' \in \partial B$ is near b, we can write $b' = b + t\dot{b} + O(t^2)$, and so

$$\begin{aligned}\int_{\partial B}\int_{\partial B} \sigma(b)\sigma(b')\psi\big(\frac{b-b'}{h}\big)h^{-2}\,db'\,db &\sim h^{-1}\int_{\partial B}\int_{-\infty}^{\infty} \sigma(b)\sigma(b)\psi(t\dot{b})\,dt\,db \\ &\sim h^{-1}C_\phi \int_{\partial B} \sigma^2(b)\,db,\end{aligned}$$

where we define

$$\int \psi(t\dot{b})\,dt = C_\phi,$$

which is independent of the choice of unit vector $\dot{b}$ because ψ is isotropic. We conclude that

$$\int (\frac{d}{dt}(\phi_h \star I_t))^2 \, dx \;\sim\; h^{-1} C_\phi \int_{\partial B} \sigma^2(b) \, db, \quad h \to 0.$$

Combining this result for both (I_t), a path of interest, and (I_t^τ), a line segment between landmark points, gives the result.

References

[1] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[2] S. Nayar, S. Baker, and H. Murase, "Parametric feature detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, California, 1996, pp. 471–477.

[3] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions?" *International Journal of Computer Vision*, vol. 28, no. 3, pp. 1–16, 1998.

[4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality analysis by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2003.

[5] D. Donoho and C. Grimes, "Hessian Eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003.

[6] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis.* Academic Press, Harcourt Brace, 1979.

[7] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling.* Chapman, Hall, 1994.

[8] C. Grimes, "New methods in nonlinear dimensionality reduction," Ph.D. dissertation, Department of Statistics, Stanford University, Stanford, CA, 2003.

[9] D. L. Donoho and C. Grimes, "When does Isomap recover the true parametrization of manifolds of articulated images?" Department of Statistics, Stanford University, Technical Report TR2002-27, 2002.

[10] A. Gray, *Tubes.* Addison-Wesley, 1990.